# Leading global retailer achieves data placement flexibility to satisfy the needs of all information consumers

## About the global retailer

This leading global retailer is a British multinational groceries and general merchandise retailer headquartered in the UK. They have revenues of over £65 Billion from a portfolio of 3,700 retail outlets and employs over 330,000 individuals.

## The organization is focused on

- Data and compute segmentation
- Retail demand forecasting
- Disaster recovery
- Financial analytics
- Data orchestration

## Initial objectives

The retailer has been a Cirata customer for over 5 years. The initial use of Cirata technology was to replicate data between their production and disaster recovery (DR) environments. The current solution enables them to support near–zero recovery time objective (RTO) and recovery point objective (RPO) in the case of an issue that results in their production environment failing. To extend the use of Cirata technology, the retailer wanted to segment a subset of their data into a separate on–premises environment to provide a smaller, more relevant and succinct data set to perform analysis and other machine learning and analytical workloads, whilst the larger data set was focused on retail demand forecasting (RDF).

## Challenges

Due to the volumes of data gathered through the retail operations of the company, there was a requirement to provide cost effective on–premises storage to manage this data, and to that end the customer chose to use IBM® Storage Ceph. This software–defined storage platform delivers a cloud–like experience while retaining the cost benefits and data sovereignty advantages of on–premises IT. To take advantage of this environment, the customer needed to be able to populate this new storage platform from their existing Hadoop clusters whilst still conducting their day–to–day business.

## Solution

The Ceph storage is Amazon Simple Storage Service (S3) compatible, and Cirata Data Migrator (CDM) moves the HDFS data into this storage platform. CDM provided the flexibility to configure migrations on a granular level to meet the retailer's specific needs for this new environment. Once the migration is started, CDM transfers the initial data sets with a single pass through the source storage, eliminating the overhead of repeated scans while also supporting continuous replication of any ongoing changes to the defined data sets. This automated process ensured the data is kept up to date, so the retailer does not need to manually refresh the new environment. This approach eliminates the risk of data loss, costly downtime and repeated attempts of data copy.

### Outcome

Data Migrator enabled the organization to:

- Continuously move data to where and when it is needed.
- Provide appropriate compute resources to the many different types of users whilst sharing the same data lake.
- Provided a standard mechanism for moving data whether it be a cloud target or on-premises.
- Future proofed the solution to take advantage of newer analytics and Metadata storage.
- To date 4PB of data moved.

## The future roadmap

The retailer continues to separate their compute away from their storage infrastructure to provide a more agile analytics capability to the business. There is also a longer-term plan to migrate fully from the legacy Hadoop infrastructure to the Ceph based architecture and Cirata technology has a large part to play in this transition.

For ultimate flexibility they intend to adopt an open data format. The first project to take advantage of this will be a new finance data hub specifically set up for the analytical requirements of the Finance business unit. The finance data hub will hold ERP data augmented with reference data from their existing clusters. This is a critical data set providing valuable information to the senior leadership team.

The open data platform will be based on the Iceberg format which is quickly becoming the de-facto standard for data storage as it allows a much wider variety of analytics and AI tools to access data without any additional data conversion. Cirata is working with the customer to develop the functionality to support their iceberg requirements.

To support the future roadmap, Cirata is working with the customer to provide a ubiquitous data movement solution to satisfy all their data movement needs, the intent is to replace other legacy technologies to establish a data orchestration center of excellence whilst providing federated autonomy across the business.

This orchestration platform will deliver:

- A single skill set for all data movement needs.
- Iceberg to Iceberg data movement.
- Live data migration for Ceph storage.
- Greater flexibility in backup, disaster recovery and test data strategies.
- Observability across all data movement.

www.cirata.com

agrcs-0425